WEEDS IDENTIFICATION USING EVOLUTIONARY ARTIFICIAL INTELLIGENCE ALGORITHM

Ahmed M. Tobal and Sahar A. Mokhtar

Department of Computers and Systems, Electronics Research Institute (ERI), Dokki, Giza, Egypt

ABSTRACT

In a world reached a population of six billion humans increasingly demand it for food, feed with a water shortage and the decline of agricultural land and the deterioration of the climate needs 1.5 billion hectares of agricultural land and in case of failure to combat pests needs about 4 billion hectares. Weeds represent 34% of the whole pests while insects, diseases and the deterioration of agricultural land present the remaining percentage. Weeds Identification has been one of the most interesting classification problems for Artificial Intelligence (AI) and image processing. The most common case is to identify weeds within the field as they reduce the productivity and harm the existing crops. Success in this area results in an increased productivity, profitability and at the same time decreases the cost of operation. On the other hand, when AI algorithms combined with appropriate imagery tools may present the right solution to the weed identification problem. In this study, we introduce an evolutionary artificial neural network to minimize the time of classification training and minimize the error through the optimization of the neuron parameters by means of a genetic algorithm. The genetic algorithm, with its global search capability, finds the optimum histogram vectors used for network training and target testing through a fitness measure that reflects the result accuracy and avoids the trial-and-error process of estimating the network inputs according to the histogram data.

1. INTRODUCTION

Weeds are unwanted plants that reduce available moisture, nutrients, sunlight and growing space needed by crop plants. Their presence can reduce crop growth, quality and yielding. In addition, they can make harvest difficult. Weeds also provide cover for diseases, insects and animals (rodents, box turtles, snakes). There are several methods that should be combined and a lot of efforts should be coordinated to control weeds including cultural, mechanical and chemical methods. Herbicides are another weed control aid that some specialists employ. Increased use of herbicides in the crop and non-crop areas, such as hedgerows, results in dramatic reduction in numbers and types of weeds. However, this causes an even greater reduction in the number of species and populations of beneficial insects (those that

kill crop pests). This resulted in an increase in pest populations, requiring more pesticides to be used, resulting in increased costs.

Modern technologies of cultivated crop growth during plant growing, which use chemical compositions to avoid weeding, lead to pollution of the environment and products of plant growing with these compositions leading to decayed production. Moreover, purchasing weed killers considerably increases the costs of the crop cultivation. That is why researchers from all over the world are trying to develop agricultural technologies of less usage of herbicides and other chemical compositions. The proposed solution is based on an idea of plant recognition with the possibility of two-classes: Useful ones and weeds. A posterior information obtained as the result of the recognition is used to realize adaptive control of the end effectors of an agricultural machine.

Corresponding Author: Ahmed M. Tobal, Department of Computers and Systems, Electronics Research Institute (ERI), Dokki, Giza, Egypt

At the chemical method of weeding, the herbicides are applied only to the plots with the weeds. This method allows not applying the herbicides to the plots with either no crop at all or where the crops are damaged (dried) or they belong to the useful species. While realization of the plant recognition with their division not into the two but into several classes it is possible to disperse different herbicides depending on sorts of the weeds. It improves the application of the herbicides to a considerable degree. Such approaches allow substantial reduction of the amount of the herbicides, which are used in crop cultivation at high probabilities of the correct recognition. The reduction of the herbicide usage and the ecological pressure to the soil accordingly can reach 50%.

Venkatesh and Thangaraj (2008) applied Artificial Intelligence (AI) techniques to find the best match of crop(s) for the given type of soil characteristics. In this paper, AI is used for weeds identification using an evolutionary artificial neural network to minimize the time of classification training and minimize the error through the optimization of the neuron parameters by means of a genetic algorithm.

The problem has two folds: First we need to represent weeds in a form suitable for identification. Then, secondly we need to identify weeds in a field and target them without affecting the crop and the existing species. For the first problem, we have used remote sensing as the tool, while AI based algorithms are used to complete the identification function (Ramli *et al.*, 2009). In our research work we have used an evolutionary artificial neural networks composed of self organizing maps network trained by genetic algorithm which is introduced and tested in the following sections.

2. DATA CAPTURING AND PROCESSING

Every system emits, absorbs, transmits or reflects electromagnetic radiation in a manner characteristic of that substance. Plant health is affected by numerous variables, most of which cause changes in the spectra of electromagnetic radiation reflected from the plant. These effects can be detected. This is the underlying principle involved in all remote sensing and on which radiometers development and manufacturing companies bases their spectral radiometers, designed for field/planets radiometry.

Using a selected number of spectral narrow-bands in the visible and near infra-red regions of the electromagnetic spectrum, gives enough essential information then it can be identified to express the type of plant and its health variables. The spectral quality of light reflected from leaves, manifested in leaf color and has long been relied upon as an indicator of plant stress as resulted in (Le Maire *et al.*, 2004; Pavan *et al.*, 2004). However, spectral characteristics of radiation reflected, transmitted, or absorbed by leaves can provide a more thorough understanding of physiological responses to growth conditions and plant adaptations to the environment.

In the late nineteenth and early twentieth centuries, technological advances began to allow the examination of changes in leaf spectra that occur with stress as shown in the results of the following references (Tellaeche *et al.*, 2011; Gardner and Blad, 1986; Tellaeche *et al.*, 2008). Investigation of such spectral characteristics has intensified greatly since the 1960s, along with the development of instrumentation and interest in the potential of remote sensing for stress detection. Largely as a result of interests in remote sensing, leaf reflectance has been studied more extensively than transmittance or absorbance responses to stress. Pioneering efforts in this field have been reviewed elsewhere in (Jacquemoud and Baret, 1990). Throughout this research history, the extent to which differing causes of stress within a species may yield correspondingly different spectral signatures has remained in question. Also in question is the degree to which the spectral response to a particular stressor may vary among species.

These changes were spectrally similar among many common stressors and vascular plant species. Increased reflectance in the far-red 690-720 nm spectrum is a particularly generic response, providing an earlier or more consistent indication of stress than reflectance in other regions of the incident solar spectrum.

It has long been suggested that alterations of reflectance in the visible spectrum by stress conditions result from the sensitivity of leaf chlorophyll concentrations to metabolic disturbance (Jacquemoud and Baret, 1990). Indeed, several studies have shown that indices based on reflectance in the far-red can precisely estimate leaf chlorophyll concentration. Thus, leaf optical properties in a relatively narrow spectral band near 700 nm are crucial for plant stress detection and the estimation of leaf chlorophyll concentration.

Spectral cube software package was used in our work for data capturing and preprocessing. It acts as data acquisition and storage software for SPECIM spectral cameras with different digital interfaces. It is a software package designed for real-time hyper spectral image

acquisition and storage with line-scan spectral imagers-SPECIM spectral cameras. In the meantime the spectral cube is mainly designed for off-line applications where the data analysis and classification is done afterwards. It also supports accessories like mechanical shutter and motion devices, e.g., SPECIM mirror scanner and x-stages of linear slides. These are required to image the samples with a line scan spectral camera. Spectral cube has all the required basic features for hyper spectral imaging like spectral calibration from calibration file, spectral band selection/reselection and binning. It also includes the basic data calibration and normalization with white and dark reference. For the visualization, it provides raw image display, false color image composition (waterfall) and the possibility to monitor measurements with spectral and spatial profiles. The data format used with SpectralCube to record the data cubes, 3D spectral images, is ENVI compatible raw bil16 format. ENVI is a hyper spectral data processing software from ITT, USA. Data can easily be imported to standard data processing software like matlab. **Figure 1** shows the data capturing setup within the lab interfaced to the used software package. The capturing mechanism is composed of computer controlled mounting body system combined with a radiometer. Leafs are mounted on the moving cart facing the radiometer with the light source and moves in steps controlled by the computer interface. In our case 512 steps with 512 snap shots captured through the radiometer resembling 512 lines for each leaf are transferred to the computer to form a single data sample. The readings received by the software expresses the leaf reflectance, which is altered by stress more consistently at visible wavelengths (400-720 nm) than in the remainder of the incident solar spectrum (730-2500 nm) (Dawson *et al.*, 1998; Rudorff and Batista, 1990).

3. MATERIALS AND METHODS

Self-organizing neural networks represent a very important class of ANNs (Holland, 1992). Such networks can learn to detect regularities and correlations in their input and adapt their future responses to that input accordingly. Self-organizing maps learn to recognize groups of similar input vectors in such a way that neurons physically near each other in the competitive layer respond to similar input vectors and automatically learns to classify them. However, the classes that the competitive layer finds are dependent only on the distance between input vectors. If two input vectors are very similar, the competitive layer probably puts them in the same class. There is no mechanism in a strictly competitive layer design to say whether or not any two input vectors are in the same class or different classes (Lihua *et al.*, 2012; Cruz-Ramírez *et al.*, 2012).

While training the neurons in a competitive layer are distributed to recognize frequently presented input vectors. The architecture for a competitive network is shown in **Fig. 2** where the input vector P and weight matrix IW1,1 are accepted to produce a vector having S1 elements which present the negative of the distances between the input vector and iIW1,1 vectors formed from the rows of the input weight matrix. The net input n1 of a competitive layer is computed by finding the negative distance between input vector p and the weight vectors and adding the biases b. If all biases are zero, the maximum net input a neuron can have is 0. This occurs when the input vector p equals that neuron's weight vector. The competitive transfer function accepts a net input vector for a layer and returns neuron outputs of 0 for all neurons except for the winner which returns output 1, the neuron associated with the most positive element of net input n1. If all biases are 0, then the neuron whose weight vector is closest to the input vector has the least negative net input and, therefore, wins the competition to output a 1.

3.1. The Training Process: The Use of Genetic Algorithms

Evolutionary algorithms are probabilistic search algorithms that simulate natural evolution. Genetic Algorithms (GAs) are one of these types of algorithms (Menon, 2004; Koza, 1990). They are based on the mechanics of natural selection and natural genetics. They combine survival of the fittest among string structures. In GAs the search space of the problem is represented as a collection of individuals. The individuals are represented by character strings, which are often referred to as chromosomes. The purpose of the use of a GA is to find the individual from the search space with the best genetic material. The quality of an individual is measured with an objective function. The part of the search space, which is to be examined, is called the population (Bharathi and Shanthi, 2012).

Roughly, a GA works in consecutive steps starting with choosing the initial population and determining the quality of each individual to select the proper parents from the population. These parents produce children, which are added to the population. For all newly created individuals a probability near zero exists that they mutate. After that, some individuals are removed from the population according to a selection criterion in order to reduce the population to its initial size. Each iteration of the algorithm is referred to as a generation and iterations stop when the algorithm output reaches predefined target with certain accuracy.

(a)   (b)
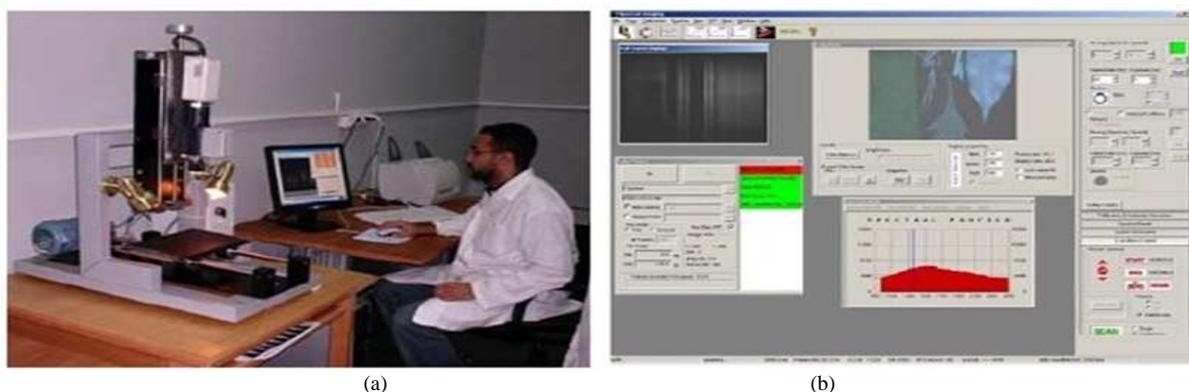
Fig. 1. Data capturing setup in the lab (a) camera with moving mechanism (b) software capturing tool

$$n_i^1 = - \|_i IW_{1,1}-p\|$$
$$a^1 = compet(n^1)$$
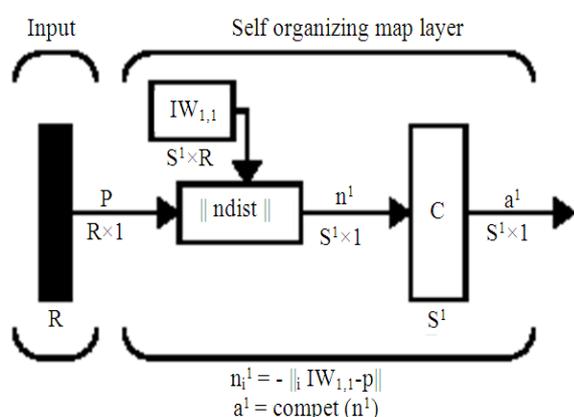
Fig. 2. SOM competitive layer

The operators which define the child production process and the mutation process are called the crossover operator and the mutation operator respectively (Rhaman and Endo, 2010; Lin *et al.*, 2011). Mutation is needed to explore new states and helps the algorithm to avoid local optima. Crossover should increase the average quality of the population. By choosing adequate crossover and mutation operators as well as an appropriate reduction mechanism, the probability that the GA results in a near-optimal solution in a reasonable number of iterations increases (Spears, 1995; Pelusi and Mascella, 2013).

3.2. Training Samples and Features

The data set used in this research work was collected using the Hamamatsu C8484-05G Spectral camera model QE and spectral cube software. The capture system sensor type is 8.67×6.60 mm -1344×1024 pixels, spectrograph model V10E 2/3", 30 µm slit size, 400-1000 nm nominal spectral range, 2.73 nm nominal spectral resolution, 0.125 nm nominal spatial bending and 2.4 as numerical aperture. The data set consists of 400 samples for four different types of crops, one hundred samples for each. Seventy five samples of each crop were used in data processing and training of the classifier system and the rest samples were used to test the performance of the trained classifier. Orange, flower, wheat and weed present the sample types used in this research to constitute two classes as useful crops and weed in one round. In other round four classes were classified to discriminate among different types of useful plants. Data captured and processed using spectral cube, then read by matlab as excel sheets, 512 values per each planet sample to build a matrix of training samples with dimensions 512×300 and testing samples matrix of size 512×100.

In our measurements, the available light source played an important role as the spectrum of the source is assumed the main reference of the measured values and dramatically affects the output resultant spectral strength distribution. We have to mention here also that the position of the camera with respect to the sun light, in day time and the time of sampling or interference from other light sources affects the accuracy of the measured values so we tried as much as we could the measuring environmental effects especially as we don't have an optimum light source characteristics which simulates the sun and covers the full range of spectrum the camera covers and complies with our range of interest. **Figure 3** shows the spectral diagrams of the training data samples as plotted using Matlab for the four types of crops used in our implementation. Each trace resembles one leaf sample, the X axis shows the no of lines scanned per sample and the Y axis demonstrates the strength calculated for each line.
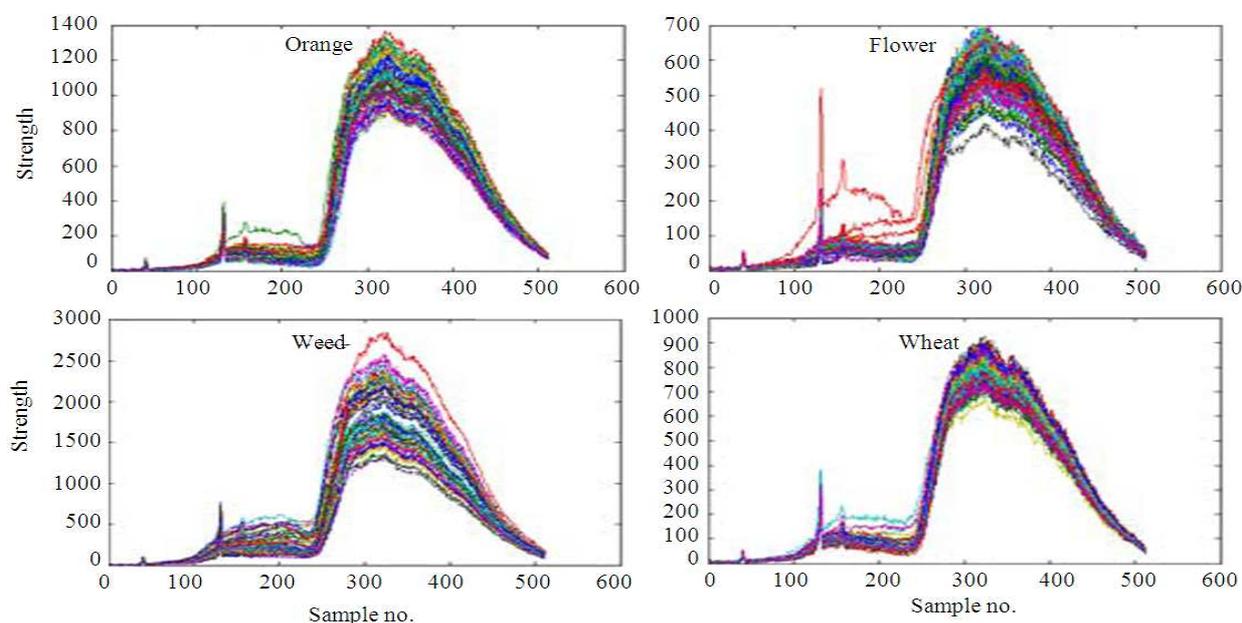
Fig. 3. Training data samples spectral diagrams

The 512 lines calculated in the 512 dots of each trace constitute one sample or one input to the classifier that will be explained in details in the following section.

4. RESULTS

The dimensions of the used hidden layer SOM network are 6X6 single layer to classify four different types of planets: Orange; flower; wheat and weed. 1000 E-books were used, 0.01 training accuracy and one tuning phase neighborhood distance. The input spectral data values selected with some statistical parameters calculated for the spectral curve to constitute 36 real values in a vector for each sample. All samples were arranged in a matrix P and used for training the SOM network.

In our application of the evolutionary algorithm, each individual, or chromosome, in the training data population specifies a set of 36 spectral values with reference values within the range 1-512 as stored in each data sample. Each spectral reference is encoded as a floating-point number and is regarded as a gene of the chromosome. The neuron evolution uses the basic strategy of GAs for evaluating and recombining the fittest individual neurons. As stated above the strategy is repeated over two phases: An evaluation phase and a reproduction phase. During the evaluation phase, the input features to the network are evaluated based on the performance of the network in which they participate. The performance is measured as the percentage of the network results exceeds according to the employed fitness function. In the reproduction phase, genetic operators that are selected by fitness rank, one-point crossovers and mutation, are used to obtain new chromosome. Every two chromosomes in the top 50% of the population (according to the fitness rank), are selected for mating. Each mating operation creates two offspring through arithmetic crossover. This process produces two complementary linear combinations of the parents X and Y as indicated in Equation 1:

$$X = rX + (1 - r)Y \text{ and } Y = (1 - r)X + rY \qquad (1)$$

where, $r = u(a_i, b_i)$ is a uniform random number between 0 and 1.

New off springs replace the bottom 50% worst performing chromosomes in the population. Finally, mutation at a rate of 1% was implemented by randomly selecting one variable and setting it equals to a uniform random number u(ai, bi), where ai and bi are the lower and upper bounds of the variable respectively. The fittest individual was not mutated, but copied to the next generation to ensure the survival of the best solution till the end of the process to build the optimized network. The implemented classifier output is demonstrated in **Table 1** for two cases: Two classes as useful crops and weed in one round and four classes to discriminate among different types of useful plants in other round.

Table 1. Experimental results

No. of classes	Parameters	SOM	Evolutionary SOM
2	No. of errors	7	2
	% success	93	98
4	No. of errors	11	3
	% success	89	97

5. DISCUSSION

Compared to the results demonstrated in Burks and Shearer research work (Burks *et al*., 2000) the achieved results in our research work showed batter performance by means of the algorithm output accuracy and the implementation simplicity in addition to the optimized time of processing and identification which paves the road for hardware implementations of such evolutionary algorithms used in real time applications. We believe that the developed mechanism controlled by a hardware embedded system based on the evolutionary algorithm when used in the field acts as online decision supporting system preventing excess use of herbicides which harm our health. According to (WRL, 2014) report concerning weeds distribution in Egypt and all over the world with a population of six billion humans increasingly demand it for food, feed with a water shortage and the decline of agricultural land and the deterioration of the climate needs 1.5 billion hectares of agricultural land and in case of failure to combat pests needs about 4 billion hectares. Weeds represent 34% of the whole pests while insects, diseases and the deterioration of agricultural land present the remaining percentage so a lot of work has to be done to secure the necessary and safe food using technology.

6. CONCLUSION

In this paper, we presented evolutionary self-organizing map neural network and demonstrated its capabilities to recognize and classify image patterns that represents different planet leaves including weeds. The intent of the evolution process is the maximization of the neural network classification performance. The network evolution uses the basic strategy of GAs for evaluating and recombining the fittest input parameters as spectral feature reference values. The strategy is repeated over two phases: An evaluation phase and a reproduction phase. During the evaluation phase, leaf optical features are evaluated based on the performance of the network in which they participate. The obtained results indicate that GA, with its global search capability, finds the optimum features. These evolved features enhanced the performance of the classification system. Also, the use of

GA eliminates the trial-and-error process of estimating the leaf features used in our experiments.

Since neural network classifiers are well suited for real-time control applications, the achieved results suggest that vision systems can achieve real-time discrimination of weed so hardware and software design must be conducted to develop an integrated real-time image processing and an evolutionary neural network classification system to be used directly in the field for real time classification and control for the fertilization machines.

7. REFERENCES

Bharathi, C.R. and V. Shanthi, 2012. An effective system for acute spotting aberration in the speech of abnormal children via artificial neural network and genetic algorithm. Am. J. Applied Sci., 9: 1561-1570. DOI: 10.3844/ajassp.2012.1561.1570

Burks, T.F., S.A. Shearer, R.S. Gates and K.D. Donohue, 2000. Back propagation neural network design and evaluation for classifying weed species using color image texture. Trans. ASAE, VOL. 43: 1029-1037.

Cruz-Ramírez, M., C. Hervás-Martínez, M. Jurado-Expósito and F. López-Granados, 2012. A multi-objective neural network based method for cover crop identification from remote sensed data. Expert Syst. Applic., 39: 10038-10048. DOI: 10.1016/j.eswa.2012.02.046

Dawson, T.P., P.J. Curran and S.E. Plummer, 1998. The biochemical decomposition of slash pine needles from reflectance spectra using neural networks. Int. J. Remote Sensing 19: 1433-1438. DOI: 10.1080/014311698215540

Gardner, B. and B. Blad, 1986. Evaluation of spectral reflectance models to estimate corn leaf area while minimizing the influence of soil background effects. J. Remote Sens. Environ., 20: 183-193. DOI: 10.1016/0034-4257(86)90022-2

Holland, J., 1992. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. Cambridge. 1st Edn., MIT Press, Cambridge, MA, ISBN-10: 9780262581110, pp: 211.

Jacquemoud, S. and F. Baret, 1990. PROSPECT: A model of leaf optical properties spectra. J. Remote Sens. Environ., 34: 75-91. DOI: 10.1016/0034-4257(90)90100-Z

Koza, J.R., 1990. Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems. A technical Report to Department of Computer Science, Institution: Stanford University.

Le Maire, G., C. François and E. Dufrêne, 2004. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. J. Remote Sens. Environ., 89: 1-28.

Lihua, F., Z. Meng, L. Zhihui, L. Hongwei and H. Tingting, 2012. Adaptive hybrid feature extraction for leaf image classification by support vector machine. J. Convergence Inform. Technol., 7: 160-166.

Lin, F., W. Zeng, J. Xiahou and Y. Jiang, 2011. Optimizing for large time delay systems by BP neural network and evolutionary algorithm improving. J. Software, 6: 2050-2055.

Menon, A., 2004. Frontiers of Evolutionary Computation. 1st Edn., Springer, Boston, ISBN-10: 1402075243, pp: 271.

Pavan, G., S. Jacquemoud, G. De Rosny, J.P. Rambaut and J.P. Frangi *et al.*, 2004. RAMIS: A new portable field radiometer to estimate leaf biochemical content. Proceeding of the 7th International Conference on Precision Agriculture and Other Precision Resources Management, Jul. 25-28, Minneapolis (Minnesota, USA).

Pelusi, D. and R. Mascella, 2013. Optimal control algorithms for second order systems. J. Comput. Sci., 9: 183-197. DOI: 10.3844/jcssp.2013.183.197

Ramli, M.S., RM.T.R. Ismail, M.A. Ahmad, S.M. Nawi and M.A.M. Hussin, 2009. Improved coupled tank liquid levels system based on swarm adaptive tuning of hybrid proportional-integral neural network controller. Am. J. Eng. Applied Sci., 2: 669-675. DOI: 10.3844/ajeassp.2009.669.675

Rhaman, K. and T. Endo, 2010. Recurrent neural network classifier for three layer conceptual network and performance evaluation. J. Comput., 5: 40-48. DOI:10.4304/jcp.5.1.40-48

Rudorff, B. and G. Batista, 1990. Spectral response of wheat and its relationship to agronomic variables in the tropical region. J. Remote Sens. Environ., 31: 53-63. DOI:10.1016/0034-4257(90)90076-X

Spears, W.M., 1995. Recombination Parameters. In: Handbook of Evolutionary Computation, IOP Publishing LTD and Oxford University Press.

Tellaeche, A., G. Pajares, X.P. Burgos-Artizzu and A. Ribeiro, 2011. A computer vision approach for weeds identification through Support Vector Machines. J. Applied Soft Computing-ASC, 11: 908-915. DOI: 10.1016/j.asoc.2010.01.011

Tellaeche, A., X.P. Burgos-artizzu, G. Pajares and A. Ribeiro, 2008. A vision-based method for weeds identification through the Bayesian decision theory. J. Pattern Recognition-PR, 41: 521-530. DOI: 10.1016/j.patcog.2007.07.007

Venkatesh, E.T. and D.P. Thangaraj, 2008. Self-organizing map and multi-layer perceptron neural network based data mining to envisage agriculture cultivation. J. Comput. Sci., 4: 494-502. DOI: 10.3844/jcssp.2008.494.502

WRL, 2014. Distribution of harmful weeds in Egypt and the world. Agricultural Research Center, Ministry of Agriculture, Weed Research Laboratory, Egypt.